

Variance of the number of Comparisons of Randomised Quicksort

Vasileios Iliopoulos and David Penman
Department of Mathematical Sciences
University of Essex, CO4 3SQ, U.K
`viliop@essex.ac.uk`, `dbpenman@essex.ac.uk`

May 2010

Contents

1	Introduction	2
2	Proof	3
3	Solution of the recurrence for B_n	6

1 Introduction

In this paper, we present a calculation of the variance of the number of comparisons required by the Quicksort algorithm for sorting a set, when the pivot is chosen uniformly and at random from the n objects $\{x_1, \dots, x_n\}$ (which have a total order on them, but not one initially known to us) to be sorted. Remember that, given a pivot x_i , the Quicksort proceeds by carrying out pairwise comparisons (which we assume can be done) of all the other objects with x_i , and using this to split the original set into two subsets, all those elements above the pivot and all those below it. We then iterate this process, choosing pivots in each smaller set uniformly at random and using comparisons with the pivot to split each set into two others. Eventually we will have all the elements in order and the algorithm terminates. The object of interest is the number C_n of comparisons required to get the n elements in order. If pivots in each set are chosen from all elements in the set uniformly at random, C_n is clearly a random variable. It is well-known that the mean M_n of C_n is equal to $2(n+1)H_n - 4n$, where $H_n = 1 + 1/2 + 1/3 + \dots + 1/n$ is the n th harmonic number. (Note that $H_0 = 0$). For proofs of this fact, see [1], [2]. We also define the n th harmonic number of order k to be equal to $H_n^{(k)} = 1 + 1/2^k + 1/3^k + \dots + 1/n^k$.

In this paper, we obtain the variance of C_n . The formula for this is stated without proof in Knuth [3], who in his Exercise 6.2.2-8 states the formula

$$\text{Var}(C_n) = 7n^2 - 4(n+1)^2 H_n^{(2)} - 2(n+1)H_n + 13n.$$

Similarly, the papers [4] and [5] provide sketches of how to prove this fact. Also, in [6] the asymptotic variance of the random variable

$$\frac{C_n - 2(n+1)(H_{n+1} - 1)}{n+1}$$

is obtained using results about moments of ‘the depth of insertion’ in a tree and some martingale arguments. However we are not aware of any source where all details of the argument are written out explicitly with as few prerequisites as possible. Thus we felt it would be desirable to provide such an account, though we freely acknowledge that not all the details of the computation are particularly interesting. No originality is claimed for the result.

The basic strategy of the argument is to use a sequence of reductions of the problem. We first use generating functions to show that it is sufficient

to prove that a certain sequence B_n , defined to the next section is equal to

$$B_n = 2(n+1)^2(H_n^2 - H_n^{(2)}) - H_n(n+1)(8n+2) + \frac{23n^2 + 17n}{2}.$$

The proof of this in turn relies on various identities involving harmonic numbers and much manipulative algebra - readers may prefer to use MAPLE at some stages (as we did ourselves to initially find the relationships, though we do include proofs for completeness).

2 Proof

The theorem that we will prove at this paper is

Theorem 2.1 *If C_n is the number of comparisons used by Quicksort with a pivot chosen uniformly at random, then*

$$\text{Var}(C_n) = 7n^2 - 4(n+1)^2 H_n^{(2)} - 2(n+1)H_n + 13n.$$

We start with a recurrence for the generating function of C_n , namely $f_n(z) = \sum_{k=0}^{n(n-1)/2} P(C_n = k)z^k$. We will use this to reduce the proof of the theorem to proving a certain recurrence formula for a quantity to be called B_n (defined below).

Theorem 2.2 *In Random Quicksort of n objects, the generating functions f_i satisfy*

$$f_n(z) = \frac{z^{n-1}}{n} \sum_{j=1}^n f_{j-1}(z) f_{n-j}(z).$$

Proof. Using the following equation

$$C_n = C_{U_n-1} + C_{n-U_n} + n - 1$$

we have that

$$\begin{aligned} P(C_n = k) &= \frac{1}{n} \sum_{m=1}^n P(C_n = k | U_n = m) \\ &= \frac{1}{n} \sum_{m=1}^n \sum_{j=1}^{k-(n-1)} P(C_{m-1} = j) P(C_{n-m} = k - (n-1) - j). \end{aligned}$$

(We are using here the fact that C_{m-1} and C_{n-m} are independent). Thus

$$P(C_n = k)z^k = \frac{1}{n} \sum_{m=1}^n \sum_{j=1}^{k-(n-1)} P(C_{m-1} = j)z^j P(C_{n-m} = k - (n-1) - j)z^{k-(n-1)-j}z^{n-1}.$$

Multiplying by z^k and summing over k , so as to get the generating function f_n of C_n on the left, we obtain

$$\begin{aligned} f_n(z) &= \frac{1}{n} \sum_{k=1}^{n-1+j} \sum_{m=1}^n \sum_{j=1}^{k-(n-1)} P(C_{m-1} = j)z^j P(C_{n-m} = k - (n-1) - j)z^{k-(n-1)-j}z^{n-1} \\ &= \frac{z^{n-1}}{n} \sum_{m=1}^n \sum_{j=1}^{k-(n-1)} P(C_{m-1} = j)z^j \sum_{k=1}^{n-1+j} P(C_{n-m} = k - (n-1) - j)z^{k-(n-1)-j}. \end{aligned}$$

Thus

$$f_n(z) = \frac{z^{n-1}}{n} \sum_{m=1}^n f_{m-1}(z)f_{n-m}(z) \quad (*)$$

as required. •

This of course leads to a recursion for the variance, using the well-known link between variance of a random variable X and its generating function $f_X(z)$:

$$\text{Var}(X) = f_X''(1) + f_X'(1) - (f_X'(1))^2.$$

We use this formula together with equation (*) above. For the first derivative,

$$f_n'(z) = \frac{(n-1)z^{n-2}}{n} \sum_{j=1}^n f_{j-1}(z)f_{n-j}(z) + \frac{z^{n-1}}{n} \sum_{j=1}^n f_{j-1}'(z)f_{n-j}(z) + \frac{z^{n-1}}{n} \sum_{j=1}^n f_{j-1}(z)f_{n-j}'(z).$$

From standard properties of generating functions, $E(C_n) = f'_n(1)$. Differentiating again we obtain

$$\begin{aligned}
f''_n(z) &= \frac{(n-1)(n-2)z^{n-3}}{n} \sum_{j=1}^n f_{j-1}(z)f_{n-j}(z) + \frac{(n-1)z^{n-2}}{n} \sum_{j=1}^n f'_{j-1}(z)f_{n-j}(z) \\
&+ \frac{(n-1)z^{n-2}}{n} \sum_{j=1}^n f_{j-1}(z)f'_{n-j}(z) + \frac{(n-1)z^{n-2}}{n} \sum_{j=1}^n f'_{j-1}(z)f'_{n-j}(z) \\
&+ \frac{z^{n-1}}{n} \sum_{j=1}^n f''_{j-1}(z)f_{n-j}(z) + \frac{z^{n-1}}{n} \sum_{j=1}^n f'_{j-1}(z)f'_{n-j}(z) + \frac{(n-1)z^{n-2}}{n} \sum_{j=1}^n f_{j-1}(z)f'_{n-j}(z) \\
&+ \frac{z^{n-1}}{n} \sum_{j=1}^n f'_{j-1}f'_{n-j}(z) + \frac{z^{n-1}}{n} \sum_{j=1}^n f_{j-1}(z)f''_{n-j}(z).
\end{aligned}$$

Setting $z = 1$, we have (see [4])

$$\begin{aligned}
f''_n(1) &= (n-1)(n-2) + \frac{2}{n}(n-1) \sum_{j=1}^n M_{j-1} + \frac{2}{n}(n-1) \sum_{j=1}^n M_{n-j} \\
&+ \frac{1}{n} \sum_{j=1}^n (f''_{j-1}(1) + f''_{n-j}(1)) + \frac{2}{n} \sum_{j=1}^n M_{j-1}M_{n-j}.
\end{aligned}$$

where M_{j-1}, M_{n-j} are $f'_{j-1}(1), f'_{n-j}(1)$, *i.e.* the mean number of comparisons to sort a set of $(j-1)$ & $(n-j)$ elements respectively. Setting $B_n = f''_n(1)/2$, we obtain

$$\begin{aligned}
2B_n &= (n-1)(n-2) + \frac{2(n-1)}{n} \sum_{j=1}^n M_{j-1} + \frac{2(n-1)}{n} \sum_{j=1}^n M_{n-j} + \frac{1}{n} \sum_{j=1}^n (2B_{j-1} + 2B_{n-j}) \\
&+ \frac{2}{n} \sum_{j=1}^n M_{j-1}M_{n-j}.
\end{aligned}$$

But now, noting that $\sum_{j=1}^n M_{j-1} = \sum_{j=1}^n M_{n-j}$, as both sums are $M_1 + \dots + M_{n-1}$ (using the fact that $M_0 = 0$), and similarly that $\sum_{j=1}^n B_{j-1} = \sum_{j=1}^n B_{n-j}$, we see this is

$$B_n = \binom{n-1}{2} + \frac{2(n-1)}{n} \sum_{j=1}^n M_{j-1} + \frac{2}{n} \sum_{j=1}^n B_{j-1} + \frac{1}{n} \sum_{j=1}^n M_{j-1}M_{n-j}. \quad (1)$$

What this argument has shown for us is the following - compare [5] where it is also shown that this recurrence has to be solved, though no details of how to solve it are given.

Theorem 2.3 *In order to prove Theorem 2.1, it is sufficient to show that the recurrence equation (1) for B_n is satisfied by [5],*

$$B_n = 2(n+1)^2 H_n^2 - (8n+2)(n+1)H_n + \frac{n(23n+17)}{2} - 2(n+1)^2 H_n^{(2)}.$$

Proof. If we get this formula, we then have

$$\begin{aligned} \text{Var}(C_n) &= f_n''(1) + f_n'(1) - (f_n'(1))^2 = 2B_n + 2(n+1)H_n - 4n - [2(n+1)H_n - 4n]^2 \\ &= 4(n+1)^2 H_n^2 - 2(8n+2)(n+1)H_n + 2\frac{n(23n+17)}{2} - 4(n+1)^2 H_n^{(2)} \\ &\quad + 2(n+1)H_n - 4n - [2(n+1)H_n - 4n]^2 \\ &= 4(n+1)^2 H_n^2 - 2(8n+2)(n+1)H_n + n(23n+17) - 4(n+1)^2 H_n^{(2)} \\ &\quad + 2(n+1)H_n - 4n - 4(n+1)^2 H_n^2 + 16n(n+1)H_n - 16n^2 \\ &= 7n^2 + 13n - 4(n+1)^2 H_n^{(2)} - 2(n+1)H_n[(8n+2) - 1 - 8n] \\ &= 7n^2 - 4(n+1)^2 H_n^{(2)} - 2(n+1)H_n + 13n \end{aligned}$$

as required. •

3 Solution of the recurrence for B_n

We have to solve the B_n recurrence. For the sum of the M_{j-1} , the expected numbers of comparisons, we have

$$\sum_{j=1}^n M_{j-1} = \sum_{j=1}^n [2jH_{j-1} - 4(j-1)] = 2 \sum_{j=1}^n jH_{j-1} - 4 \sum_{j=1}^n (j-1).$$

For the computation of the first sum, a Lemma follows.

Lemma 3.1 *For $n \in \mathbf{N}$*

$$\sum_{j=1}^n jH_{j-1} = \frac{n(n+1)H_{n+1}}{2} - \frac{n(n+5)}{4} \text{ and } \sum_{j=1}^n M_{j-1} = n(n+1)H_{n+1} - \frac{5n^2 + n}{2}$$

Proof. By induction on n , the case $n = 1$ being trivial. Suppose that it holds for all $n \leq k$. Then for $n = k + 1$ we have

$$\begin{aligned}
\sum_{j=1}^{k+1} jH_{j-1} &= \sum_{j=1}^k jH_{j-1} + (k+1)H_k = \frac{k(k+1)H_{k+1}}{2} - \frac{k(k+5)}{4} + (k+1)H_k \\
&= \frac{k(k+1)H_{k+1}}{2} - \frac{k(k+5)}{4} + (k+1)\left(H_{k+1} - \frac{1}{k+1}\right) \\
&= \frac{(k+1)(k+2)H_{k+1}}{2} - \frac{k(k+5)}{4} - 1 \\
&= \frac{(k+1)(k+2)H_{k+2}}{2} - \frac{k+1}{2} - \frac{k(k+5)}{4} - 1 \\
&= \frac{(k+1)(k+2)H_{k+2}}{2} - \frac{k^2 + 7k + 6}{4} \\
&= \frac{(k+1)(k+2)H_{k+2}}{2} - \frac{(k+1)(k+6)}{4} \\
&= \frac{n(n+1)H_{n+1}}{2} - \frac{n(n+5)}{4}.
\end{aligned}$$

giving the first claim. The second claim follows recalling that $\sum_{j=1}^n (j-1) = n(n-1)/2$. •

Now, we will compute the term

$$\sum_{j=1}^n M_{j-1}M_{n-j}. \quad (2)$$

We shall use three Lemmas in the proof.

Lemma 3.2 *For $n \in \mathbf{N}$, it holds that*

$$\sum_{j=1}^n M_{j-1}M_{n-j} = 4 \sum_{j=1}^n jH_{j-1}(n-j+1)H_{n-j} - \frac{8}{3}n(n^2-1)H_{n+1} + \frac{44n}{9}(n^2-1)$$

Proof. To do this, we will again use the formula obtained previously for M_j . We have

$$\begin{aligned}
\sum_{j=1}^n M_{j-1} M_{n-j} &= \sum_{j=1}^n [(2jH_{j-1} - 4j + 4)(2(n-j+1)H_{n-j} - 4n + 4j)] \\
&= 4 \sum_{j=1}^n jH_{j-1}(n-j+1)H_{n-j} - 8n \sum_{j=1}^n jH_{j-1} + 8 \sum_{j=1}^n j^2 H_{j-1} \\
&\quad - 8 \sum_{j=1}^n j(n-j+1)H_{n-j} + 16n \sum_{j=1}^n j - 16 \sum_{j=1}^n j^2 + 8 \sum_{j=1}^n (n-j+1)H_{n-j} \\
&\quad - 16n^2 + 16 \sum_{j=1}^n j.
\end{aligned}$$

We need to work out the value of $\sum_{j=1}^n j^2 H_{j-1}$. Using MAPLE initially, we found

$$\sum_{j=1}^n j^2 H_{j-1} = \frac{6n(n+1)(2n+1)H_{n+1} - n(n+1)(4n+23)}{36};$$

we will confirm this by induction.

Lemma 3.3 *For $n \in \mathbb{N}$ holds*

$$\sum_{j=1}^n j^2 H_{j-1} = \frac{6n(n+1)(2n+1)H_{n+1} - n(n+1)(4n+23)}{36}$$

Proof. By induction on n , the case $n = 1$ yielding $1^2 H_0 = 0$ on the left-hand side and on the right-hand side

$$\frac{36H_2 - 54}{36} = \frac{36 + 18 - 54}{36} = 0.$$

Suppose that the equation holds for all $n \leq k$. For $n = k + 1$, we have

$$\begin{aligned}
\sum_{j=1}^{k+1} j^2 H_{j-1} &= \sum_{j=1}^k j^2 H_{j-1} + (k+1)^2 H_k \\
&= \frac{6k(k+1)(2k+1)H_{k+1} - k(k+1)(4k+23)}{36} + (k+1)^2 H_{k+1} - (k+1) \\
&= \frac{6(k+1)H_{k+1}(2k^2+7k+6) - k(k+1)(4k+23) - 36(k+1)}{36} \\
&= \frac{6(k+1)H_{k+1}(k+2)(2k+3) - (k+1)(4k^2+23k+36)}{36} \\
&= \frac{6(k+1)H_{k+2}(k+2)(2k+3) - 6(k+1)(2k+3) - (k+1)(4k^2+23k+36)}{36} \\
&= \frac{6(k+1)(k+2)H_{k+2}(2k+3) - (k+1)(4k^2+35k+54)}{36} \\
&= \frac{6(k+1)(k+2)H_{k+2}(2k+3) - (k+1)(k+2)(4k+27)}{36}
\end{aligned}$$

finishing the proof of Lemma 3.3. •

We also need to compute $\sum_{j=1}^n j(n-j+1)H_{n-j}$. We have

Lemma 3.4 For $n \in \mathbf{N}$

$$\sum_{j=1}^n j(n-j+1)H_{n-j} = \frac{6nH_{n+1}(n^2+3n+2) - 5n^3 - 27n^2 - 22n}{36}.$$

Proof. We can write $j = n+1 - (n-j+1)$. Then, substituting $k = n-j+1$ we obtain

$$\begin{aligned}
\sum_{j=1}^n j(n-j+1)H_{n-j} &= \sum_{j=1}^n [(n+1) - (n-j+1)](n-j+1)H_{n-j} \\
&= (n+1) \sum_{j=1}^n (n-j+1)H_{n-j} - \sum_{j=1}^n (n-j+1)^2 H_{n-j} \\
&= (n+1) \sum_{k=1}^n kH_{k-1} - \sum_{k=1}^n k^2 H_{k-1}.
\end{aligned}$$

Thus, since we know both sums by Lemmas 3.1 and 3.3 we get

$$\begin{aligned}
\sum_{j=1}^n j(n-j+1)H_{n-j} &= (n+1) \sum_{k=1}^n kH_{k-1} - \sum_{k=1}^n k^2 H_{k-1} \\
&= (n+1) \left(\frac{n(n+1)H_{n+1}}{2} - \frac{n(n+5)}{4} \right) \\
&\quad - \frac{6n(n+1)(2n+1)H_{n+1} - n(n+1)(4n+23)}{36} \\
&= \frac{n(n+1)^2 H_{n+1}}{2} - \frac{n(n+1)(n+5)}{4} \\
&\quad - \frac{6n(n+1)(2n+1)H_{n+1} - n(n+1)(4n+23)}{36} \\
&= \frac{18n(n+1)^2 H_{n+1} - 9n(n+1)(n+5)}{36} \\
&\quad - \frac{6n(n+1)(2n+1)H_{n+1} - n(n+1)(4n+23)}{36} \\
&= \frac{6nH_{n+1}(n^2 + 3n + 2) - n(n+1)(5n+22)}{36}
\end{aligned}$$

which is easily checked to be equal to the quantity in the statement above on expanding out. •

We are now ready to complete the evaluation of $\sum_{j=1}^n M_{j-1}M_{n-j}$. Note first that $\sum_{j=1}^n (n-j+1)H_{n-j} = \sum_{k=1}^n kH_{k-1}$ (set $k = n-j+1$) and so

Lemma 3.1 can be used to compute it. Pulling everything together, we have

$$\begin{aligned}
\sum_{j=1}^n M_{j-1} M_{n-j} &= 4 \sum_{j=1}^n j H_{j-1} (n-j+1) H_{n-j} - 8n \left(\frac{n(n+1)H_{n+1}}{2} - \frac{n(n+5)}{4} \right) \\
&\quad + 8 \left(\frac{6n(n+1)(2n+1)H_{n+1} - n(n+1)(4n+23)}{36} \right) + 16n \sum_{j=1}^n j \\
&\quad - 8 \left(\frac{6nH_{n+1}(n^2+3n+2) - 5n^3 - 27n^2 - 22n}{36} \right) - 16 \sum_{j=1}^n j^2 \\
&\quad + 8 \left(\frac{n(n+1)H_{n+1}}{2} - \frac{n(n+5)}{4} \right) - 16n^2 + 16 \sum_{j=1}^n j \\
&= 4 \sum_{j=1}^n j H_{j-1} (n-j+1) H_{n-j} - 4n^2(n+1)H_{n+1} + 2n^2(n+5) \\
&\quad + \frac{8}{36} [6n(n^2-1)H_{n+1} + n^3 - n] + 8n^2(n+1) - \frac{16n^3}{3} - 8n^2 - \frac{16n}{6} \\
&\quad + 4n(n+1)H_{n+1} - 2n(n+5) - 16n^2 + 8n(n+1) \\
&= 4 \sum_{j=1}^n j H_{j-1} (n-j+1) H_{n-j} - 4n(n+1)(n-1)H_{n+1} \\
&\quad + \frac{4}{3}n(n^2-1)H_{n+1} + \frac{176n^3}{36} - \frac{176n}{36}.
\end{aligned}$$

Thus we indeed get the conclusion of Lemma 3.2, namely that

$$\sum_{j=1}^n M_{j-1} M_{n-j} = 4 \sum_{j=1}^n j H_{j-1} (n-j+1) H_{n-j} - \frac{8n}{3}(n^2-1)H_{n+1} + \frac{44n}{9}(n^2-1).$$

Returning back to the recurrence relation (1), we obtain from Lemmas 3.1 and 3.2 that

$$\begin{aligned}
B_n &= \frac{(n-1)(n-2)}{2} + \frac{2(n-1)}{n}(n(n+1)H_{n+1} - \frac{5n^2+n}{2}) \\
&\quad + \frac{4 \sum_{j=1}^n j H_{j-1}(n-j+1)H_{n-j}}{n} - \frac{8}{3}(n^2-1)H_{n+1} + \frac{44}{9}(n^2-1) + \frac{2}{n} \sum_{j=1}^n B_{j-1} \\
&= \frac{(n-1)(n-2)}{2} + 2(n-1)(n+1)H_{n+1} - (n-1)(5n+1) \\
&\quad + \frac{4 \sum_{j=1}^n j H_{j-1}(n-j+1)H_{n-j}}{n} - \frac{8}{3}(n^2-1)H_{n+1} + \frac{44}{9}(n^2-1) + \frac{2}{n} \sum_{j=1}^n B_{j-1}.
\end{aligned}$$

Finally,

$$\begin{aligned}
B_n &= \frac{4 \sum_{j=1}^n j H_{j-1}(n-j+1)H_{n-j}}{n} + \frac{2}{n} \sum_{j=1}^n B_{j-1} + \frac{-9n^2+5n+4}{2} \\
&\quad - \frac{2}{3}(n^2-1)H_{n+1} + \frac{44}{9}(n^2-1).
\end{aligned}$$

Multiplying by n , we have

$$\begin{aligned}
nB_n &= 4 \sum_{j=1}^n j H_{j-1}(n-j+1)H_{n-j} + 2 \sum_{j=1}^n B_{j-1} + n \frac{-9n^2+5n+4}{2} \\
&\quad - n \frac{2}{3}(n^2-1)H_{n+1} + n \frac{44}{9}(n^2-1).
\end{aligned}$$

For $n+1$, we have similarly

$$\begin{aligned}
&(n+1)B_{n+1} \\
&= 4 \sum_{j=1}^{n+1} j H_{j-1}(n-j+2)H_{n+1-j} + 2 \sum_{j=1}^{n+1} B_{j-1} + (n+1) \frac{-9(n+1)^2+5(n+1)+4}{2} \\
&\quad - (n+1) \frac{2}{3}[(n+1)^2-1]H_{n+2} + (n+1) \frac{44}{9}[(n+1)^2-1].
\end{aligned}$$

Subtracting nB_n from $(n+1)B_{n+1}$, we obtain

$$\begin{aligned}
& (n+1)B_{n+1} - nB_n \\
&= 4\left[\sum_{j=1}^{n+1} jH_{j-1}(n-j+2)H_{n+1-j} - \sum_{j=1}^n jH_{j-1}(n-j+1)H_{n-j}\right] \\
&+ 2B_n + (n+1)\frac{-9(n+1)^2 + 5(n+1) + 4}{2} - n\frac{-9n^2 + 5n + 4}{2} \\
&- (n+1)\frac{2}{3}[(n+1)^2 - 1]H_{n+2} + \frac{2}{3}n(n^2 - 1)H_{n+1} \\
&+ (n+1)\frac{44}{9}[(n+1)^2 - 1] - n\frac{44}{9}(n^2 - 1) \\
&= 4\left[\sum_{j=1}^n jH_{j-1}(n-j+2)H_{n+1-j} - \sum_{j=1}^n jH_{j-1}(n-j+1)H_{n-j}\right] \\
&+ 2B_n - \frac{27n^2 + 17n}{2} - \frac{2}{3}n(n+1)(n+2)H_{n+2} + \frac{2}{3}nH_{n+1}(n^2 - 1) + \frac{44n(n+1)}{3}
\end{aligned}$$

noting that the term for $j = n+1$ does not contribute to the sum. In the first sum, we use the facts that $H_{n+1-j} = H_{n-j} + 1/(n+1-j)$ and that $n-j+2 = (n-j+1) + 1$, and then we get

$$\begin{aligned}
& 4\left[\sum_{j=1}^n jH_{j-1}(n-j+1)H_{n-j} + \sum_{j=1}^n \frac{jH_{j-1}(n-j+1)}{n-j+1}\right. \\
&+ \sum_{j=1}^n jH_{j-1}H_{n-j+1} - \sum_{j=1}^n jH_{j-1}(n-j+1)H_{n-j}] + 2B_n \\
&- \frac{27n^2 + 17n}{2} - \frac{2}{3}n(n+1)(n+2)H_{n+2} + \frac{2}{3}nH_{n+1}(n^2 - 1) + \frac{44n(n+1)}{3}.
\end{aligned}$$

The first sum on the first line cancels with the equal sum on the second line, the second sum on the first line simplifies, and again using $H_{n+2} = H_{n+1} + 1/(n+2)$ on the last line, we obtain

$$4\left[\sum_{j=1}^n jH_{j-1} + \sum_{j=1}^n jH_{j-1}H_{n-j+1}\right] + 2B_n - 2n(n+1)H_{n+1} + \frac{1}{2}n(n+11).$$

We thus see that we have to work out the following expression:

$$\sum_{j=1}^n jH_{j-1}H_{n+1-j}.$$

We note that

$$\begin{aligned}
\sum_{j=1}^n j H_{j-1} H_{n+1-j} &= 1H_0H_n + 2H_1H_{n-1} + 3H_2H_{n-2} + \dots + (n-1)H_{n-2}H_2 + nH_{n-1}H_1 \\
&= \frac{n+2}{2}(H_1H_{n-1} + H_2H_{n-2} + \dots + H_{n-2}H_2 + H_{n-1}H_1 + H_nH_0) \\
&= \frac{n+2}{2} \sum_{j=1}^n H_j H_{n-j}
\end{aligned}$$

so it suffices now to obtain the quantity $\sum_{j=1}^n H_j H_{n-j}$.

Sedgewick [7] presents and proves the following result:

Lemma 3.5

$$\sum_{i=1}^n H_i H_{n+1-i} = (n+2)(H_{n+1}^2 - H_{n+1}^{(2)}) - 2(n+1)(H_{n+1} - 1).$$

Proof.

$$\sum_{i=1}^n H_i H_{n+1-i} = \sum_{i=1}^n H_i \sum_{j=1}^{n+1-i} \frac{1}{j}.$$

The set $\{(i, j) : 1 \leq i \leq n, 1 \leq j \leq n+1-i\}$ is, as a picture easily shows, the same as $\{(i, j) : 1 \leq i \leq n+1-j, 1 \leq j \leq n\}$. Thus the above is

$$\sum_{j=1}^n \frac{1}{j} \sum_{i=1}^{n+1-j} H_i = \sum_{j=1}^n \frac{1}{j} [(n+2-j)H_{n+1-j} - (n+1-j)]$$

To see the claim about the sum of the H_j s, we note that

$$\begin{aligned}
\sum_{j=1}^n H_j &= H_1 + H_2 + \dots + H_n = 1 + (1 + \frac{1}{2}) + \dots + (1 + \frac{1}{2} + \dots + \frac{1}{n}) \\
&= n + (n-1)\frac{1}{2} + \dots + [n - (n-1)]\frac{1}{n} = n(1 + \frac{1}{2} + \dots + \frac{1}{n}) - (\frac{1}{2} + \frac{2}{3} + \dots + \frac{n-1}{n}) \\
&= n(1 + \frac{1}{2} + \dots + \frac{1}{n}) - [(1 - \frac{1}{2}) + (1 - \frac{1}{3}) + \dots + (1 - \frac{1}{n})] \\
&= nH_n - n + H_n = (n+1)H_n - n
\end{aligned}$$

Thus we get, reusing the result about $\sum_{j=1}^n H_j$,

$$\begin{aligned}
& (n+2) \sum_{j=1}^n \frac{H_{n+1-j}}{j} - \sum_{j=1}^n H_{n+1-j} - (n+1)H_n + n \\
&= (n+2) \sum_{j=1}^n \frac{H_{n+1-j}}{j} - [(n+1)H_n - n] - (n+1)H_n + n \\
&= (n+2) \sum_{j=1}^n \frac{H_{n+1-j}}{j} - 2(n+1)(H_{n+1} - 1).
\end{aligned}$$

To analyse the first sum above, we note (again following [7] here)

$$\begin{aligned}
\sum_{j=1}^n \frac{H_{n+1-j}}{j} &= \sum_{j=1}^n \frac{H_{n-j}}{j} + \sum_{j=1}^n \frac{1}{j(n+1-j)} \\
&= \sum_{j=1}^{n-1} \frac{H_{n-j}}{j} + \frac{1}{n+1} \sum_{j=1}^n \left(\frac{1}{j} + \frac{1}{n+1-j} \right)
\end{aligned}$$

and this gives that

$$\sum_{j=1}^n \frac{H_{n+1-j}}{j} = \sum_{j=1}^{n-1} \frac{H_{n-j}}{j} + 2 \frac{H_n}{n+1}. \quad (3)$$

Iterating this equation, and using $H_0 = 0$ at the end, we obtain the identity

$$\sum_{j=1}^n \frac{H_{n+1-j}}{j} = 2 \sum_{k=1}^n \frac{H_k}{k+1}. \quad (4)$$

The right-hand side is

$$\begin{aligned}
2 \sum_{k=1}^n \frac{H_k}{k+1} &= 2 \sum_{k=2}^{n+1} \frac{H_{k-1}}{k} = 2 \sum_{k=1}^{n+1} \frac{H_{k-1}}{k} = 2 \sum_{k=1}^{n+1} \frac{H_k}{k} - 2 \sum_{k=1}^{n+1} \frac{1}{k^2} \\
&= 2 \sum_{k=1}^{n+1} \sum_{j=1}^k \frac{1}{jk} - 2H_{n+1}^{(2)} = 2 \sum_{j=1}^{n+1} \sum_{k=j}^{n+1} \frac{1}{jk} - 2H_{n+1}^{(2)} = 2 \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \frac{1}{kj} - 2H_{n+1}^{(2)}.
\end{aligned}$$

Again noting that $\{(j, k) : k \leq j \leq n+1, 1 \leq k \leq n+1\}$ gives the same values of $1/(jk)$ as $\{(j, k) : 1 \leq j \leq k, 1 \leq k \leq n+1\}$ provided we note that

the terms for $j = k$ are repeated, we get

$$\begin{aligned} 2 \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \frac{1}{kj} - 2H_{n+1}^{(2)} &= \sum_{k=1}^{n+1} \left(\sum_{j=1}^{n+1} \frac{1}{kj} + \frac{1}{k^2} \right) - 2H_{n+1}^{(2)} \\ &= H_{n+1}^2 + H_{n+1}^{(2)} - 2H_{n+1}^{(2)} = H_{n+1}^2 - H_{n+1}^{(2)}. \end{aligned}$$

Thus, we have, as in the statement of Lemma 3.5,

$$\sum_{i=1}^n H_i H_{n+1-i} = (n+2)(H_{n+1}^2 - H_{n+1}^{(2)}) - 2(n+1)(H_{n+1} - 1).$$

Also, the following Corollary is obtained, using equations from the last Lemma, by Sedgewick [7].

Corollary 3.6 *For $n \in \mathbf{N}$, it holds*

$$H_{n+1}^2 - H_{n+1}^{(2)} = 2 \sum_{j=1}^n \frac{H_j}{j+1}$$

Proof. From equations (3) and (4), we see that

$$\begin{aligned} H_{n+1}^2 - H_{n+1}^{(2)} &= \sum_{j=1}^n \frac{H_{n+1-j}}{j} = \sum_{j=1}^{n-1} \frac{H_{n-j}}{j} + 2 \frac{H_n}{n+1} \\ \Rightarrow H_{n+1}^2 - H_{n+1}^{(2)} &= H_n^2 - H_n^{(2)} + 2 \frac{H_n}{n+1} \\ \Rightarrow H_{n+1}^2 - H_{n+1}^{(2)} &= 2 \sum_{j=1}^n \frac{H_j}{j+1} \end{aligned}$$

by iteration. •

We will use the above Lemma and Corollary in our analysis. We have that

$$\sum_{i=1}^n H_i H_{n+1-i} = \sum_{i=1}^n \left[H_i \left(H_{n-i} + \frac{1}{n+1-i} \right) \right] = \sum_{i=1}^n H_i H_{n-i} + \sum_{i=1}^n \frac{H_i}{n+1-i}$$

The second sum substituting $j = n+1-i$ becomes

$$\sum_{i=1}^n \frac{H_i}{n+1-i} = \sum_{j=1}^n \frac{H_{n+1-j}}{j}$$

As we have seen this is equal to

$$\sum_{j=1}^n \frac{H_{n+1-j}}{j} = H_{n+1}^2 - H_{n+1}^{(2)}$$

Hence, by Lemma 3.5

$$\begin{aligned} \sum_{i=1}^n H_i H_{n-i} &= (n+2)(H_{n+1}^2 - H_{n+1}^{(2)}) - 2(n+1)(H_{n+1} - 1) - (H_{n+1}^2 - H_{n+1}^{(2)}) \\ &= (n+1)[(H_{n+1}^2 - H_{n+1}^{(2)}) - 2(H_{n+1} - 1)]. \end{aligned}$$

Using the above equation and the result obtained in page 13, just before Lemma 3.5, we deduce that

$$\begin{aligned} \sum_{j=1}^n j H_{j-1} H_{n+1-j} &= \frac{n+2}{2}(n+1)[(H_{n+1}^2 - H_{n+1}^{(2)}) - 2(H_{n+1} - 1)] \\ &= \binom{n+2}{2}[(H_{n+1}^2 - H_{n+1}^{(2)}) - 2(H_{n+1} - 1)]. \end{aligned}$$

Having worked out all the expressions involved in the following relation, we can now finish off:

$$(n+1)B_{n+1} - nB_n = 4\left(\sum_{j=1}^n j H_{j-1} + \sum_{j=1}^n j H_{j-1} H_{n-j+1}\right) + 2B_n - 2n(n+1)H_{n+1} + \frac{1}{2}n(n+11).$$

We have

$$\begin{aligned} &(n+1)B_{n+1} - nB_n \\ &= 4\left[\frac{n(n+1)H_{n+1}}{2} - \frac{n(n+5)}{4} + \binom{n+2}{2}[(H_{n+1}^2 - H_{n+1}^{(2)}) - 2(H_{n+1} - 1)]\right] \\ &\quad + 2B_n - 2n(n+1)H_{n+1} + \frac{1}{2}n(n+11) \\ &= 2n(n+1)H_{n+1} - n(n+5) + 2(n+1)(n+2)(H_{n+1}^2 - H_{n+1}^{(2)}) \\ &\quad - 4(n+1)(n+2)(H_{n+1} - 1) + 2B_n - 2n(n+1)H_{n+1} + \frac{1}{2}n(n+11) \\ &= 2(n+1)(n+2)(H_{n+1}^2 - H_{n+1}^{(2)}) - 4(n+1)(n+2)(H_{n+1} - 1) - \frac{n(n-1)}{2} + 2B_n. \end{aligned}$$

Then,

$$B_{n+1} = 2(n+2)(H_{n+1}^2 - H_{n+1}^{(2)}) - 4(n+2)(H_{n+1} - 1) - \frac{n(n-1)}{2(n+1)} + \frac{n+2}{n+1}B_n$$

$$\frac{B_{n+1}}{n+2} = \frac{B_n}{n+1} + 2(H_{n+1}^2 - H_{n+1}^{(2)}) - 4(H_{n+1} - 1) - \frac{n(n-1)}{2(n+1)(n+2)}.$$

The last equation is equivalent to

$$\frac{B_n}{n+1} = \frac{B_{n-1}}{n} + 2(H_n^2 - H_n^{(2)}) - 4(H_n - 1) - \frac{(n-1)(n-2)}{2n(n+1)}.$$

Iterating the recurrence relation, we obtain

$$\frac{B_n}{n+1} = \frac{B_0}{1} + 2 \sum_{i=1}^n (H_i^2 - H_i^{(2)}) - 4 \sum_{i=1}^n (H_i - 1) - \sum_{i=1}^n \frac{(i-1)(i-2)}{2i(i+1)}.$$

Since $B_0 = 0$, it is

$$\frac{B_n}{n+1} = 2 \sum_{i=1}^n (H_i^2 - H_i^{(2)}) - 4 \sum_{i=1}^n (H_i - 1) - \sum_{i=1}^n \frac{(i-1)(i-2)}{2i(i+1)}.$$

The first sum, by Corollary 3.6 is equal to

$$\begin{aligned} \sum_{i=1}^n (H_i^2 - H_i^{(2)}) &= (H_1^2 - H_1^{(2)}) + (H_2^2 - H_2^{(2)}) + \dots + (H_n^2 - H_n^{(2)}) \\ &= n(H_n^2 - H_n^{(2)}) - 2 \sum_{i=1}^{n-1} \frac{iH_i}{i+1} \\ &= (n+1)(H_n^2 - H_n^{(2)}) - 2 \sum_{i=1}^{n-1} \frac{iH_i}{i+1} - 2 \sum_{i=1}^{n-1} \frac{H_i}{i+1} \\ &= (n+1)(H_n^2 - H_n^{(2)}) - 2 \sum_{i=1}^{n-1} H_i \\ &= (n+1)(H_n^2 - H_n^{(2)}) + 2n - 2nH_n. \end{aligned}$$

Note that on the third line we add and subtract simultaneously $(H_n^2 - H_n^{(2)})$, which is equal to $2 \sum_{i=1}^{n-1} H_i/i + 1$ by Corollary 3.6. Doing so, the fraction is

cancelled and the corresponding sum can be easily computed. Hence

$$\begin{aligned}
\frac{B_n}{n+1} &= 2(n+1)(H_n^2 - H_n^{(2)}) + 4n - 4nH_n - 4[(n+1)H_n - 2n] - \sum_{i=1}^n \frac{(i-1)(i-2)}{2i(i+1)} \\
&= 2(n+1)(H_n^2 - H_n^{(2)}) + 4n - 4nH_n - 4[(n+1)H_n - 2n] - \sum_{i=1}^n \left(\frac{i+2}{2i} - \frac{3}{i+1} \right) \\
&= 2(n+1)(H_n^2 - H_n^{(2)}) + 12n - 8nH_n - 4H_n - \sum_{i=1}^n \frac{i+2}{2i} + \sum_{i=1}^n \frac{3}{i+1} \\
&= 2(n+1)(H_n^2 - H_n^{(2)}) + 12n - 8nH_n - 4H_n - \frac{n}{2} - H_n + 3H_{n+1} - 3 \\
&= 2(n+1)(H_n^2 - H_n^{(2)}) + 12n - 8nH_n - 4H_n - \frac{n}{2} - H_n + 3H_n + \frac{3}{n+1} - 3 \\
&= 2(n+1)(H_n^2 - H_n^{(2)}) - H_n(8n+2) + \frac{23n}{2} + \frac{3}{n+1} - 3 \\
&= 2(n+1)(H_n^2 - H_n^{(2)}) - H_n(8n+2) + \frac{23n^2 + 17n}{2(n+1)}.
\end{aligned}$$

Finally, multiplying both sides by $n+1$ we obtain

$$B_n = 2(n+1)^2(H_n^2 - H_n^{(2)}) - H_n(n+1)(8n+2) + \frac{23n^2 + 17n}{2}.$$

Now, the Proof of Theorem 2.3 is complete. Consequently, the Variance of the number of pairwise comparisons C_n of Randomised Quicksort is equal to

$$7n^2 - 4(n+1)^2 H_n^{(2)} - 2(n+1)H_n + 13n.$$

References

- [1] *Probability and Computing: Randomized Algorithms and Probabilistic Analysis.*
Michael Mitzenmacher and Eli Upfal. Cambridge University Press, 2005.
- [2] *The number of bit comparisons used by Quicksort: An average case analysis.*
James Allen Fill, Department of Mathematical Sciences, The Johns Hopkins University. Svante Janson, Department of Mathematics, Uppsala University, 2004.
- [3] *The Art of Computer Programming. Vol. 3: Sorting and Searching.*
Donald E. Knuth, Stanford University. 2nd edition, Addison-Wesley Publishing Company, 1998.
- [4] *Probabilistic Analysis of Time Complexity of Quicksort.*
Tadashi Mizoi, Shunji Osaki. Faculty of Engineering, Hiroshima University, 1996.
- [5] *Quicksort algorithm again revisited.*
Charles Knessl, Department of Mathematics Statistics and Computer Science, University of Illinois. Wojciech Szpankowski, Department of Computer Science, Purdue University, 1999.
- [6] *A limiting distribution for Quicksort.*
Mireille Regnier, RAIRO Theoretical Informatics and Applications, 1989.
- [7] *Quicksort.*
Robert Sedgewick. Garland Publishing, New York, 1980.